A quantitative study of the benefits of co-regulation using the *spoIIA* operon as an example

Dagmar Iber

Mathematical Institute, Centre for Mathematical Biology

24-29 St. Giles, Oxford OX1 3LB, UK

Tel.: +44-1865-283875, Fax: +44-1865-273583, iber@maths.ox.ac.uk

The distribution of most genes is not random, and functionally linked genes are often found in clusters. Several theories have been put forward to explain the emergence and persistence of operons in bacteria [1]. Careful analysis of genomic data favours the co-regulation model [2, 3], where gene organization into operons is driven by the benefits of coordinated gene expression and regulation. Direct evidence that co-expression increases the individual's fitness enough to ensure operon formation and maintenance is, however, still lacking. Here, a previously described quantitative model of the network that controls the transcription factor σ^F during sporulation in *Bacillus subtilis* [4] is employed to quantify the benefits arising from both organisation of the sporulation genes into the *spoIIA* operon and from translational coupling. The analysis shows that operon organization, together with translational coupling, is important because of the inherent stochastic nature of gene expression which skews the ratios between protein concentrations in the absence of co-regulation. The predicted impact of different forms of gene regulation on fitness and survival agrees quantitatively with published sporulation efficiencies.

Running Title A quantitative study of the benefits of co-regulation
Keywords coupled gene expression/noise/operon/signaling networks/sporulation/
Subject category Chromatin & Transcription/Metabolic and regulatory networks
Characters 19650

standfirst text

The benefits of co-regulated gene expression have been suggested to drive operon emergence and persistence but direct evidence that co-expression increases an individual's fitness is lacking. Here, a previously described quantitative model of the σ^F signaling network is employed to show that the inherent noise in gene expression can be sufficiently harmful that co-regulated expression can substantial increase survival chances.

Main findings of the study

- the study provides further support for the co-regulation model for operon formation
- the study reveals that small variations in gene expression, as arise from the inherent stochasticity of biological processes, can be harmful, and that co-regulation of the expression of interacting proteins by organization of the genes into operons can substantially increase survival chances
- the quantification of the impact of co-regulation on an individual's fitness is possible for the first time because of the detailed mathematical model that we have developed recently for the genes encoded in the spoIIA operon.

1 Introduction

The available genome sequences demonstrate that many genes are clustered on chromosomes according to their function. Genes in bacteria are clustered but can also be organized into operons such that the expression of a group of genes is regulated by the same genetic control element. When operons were first discovered it was assumed that the benefit of co-transcription led to operon assembly [5]. Other models have since been proposed, and these belong to one of three classes, the natal model, the Fisher model, or the selfish operon model [1]. According to the natal model, clustering of genes is the consequence of gene duplication. However, since operons comprise genes that belong to very distant families and the majority of paralogs do not cluster, this model is insufficient to explain operon origin [1, 6]. A recast of the Fisher model, adapted to prokaryotes, proposes that clustering of genes reduces the likelihood that co-adapted genes become separated by recombination. However, this does not explain how operons can emerge, as recombination is as likely to generate clusters as to disrupt them. According to the selfish operon model, operons facilitate the horizontal transfer of functionally related genes [7]. The physical proximity of genes thus does not necessarily provide a selective advantage to the individual organism but rather to the gene cluster itself, because it can be efficiently transmitted both horizontally as well as vertically. Recent studies have, however, failed to observe the gene cluster pattern predicted by the model, and this strongly suggests that the selfish operon model does not explain the emergence and persistence of operons [2, 3]. So what drives operon assembly?

The idea that co-transcription of genes provides a selective advantage to the individual organism has never been contradicted. It has been questioned only because it remains unclear whether the benefits of co-transcription could be strong enough to drive the assembly of operons by rare recombination events [7, 1]. A genotype that confers higher fitness will dominate in a population with bounded total population size only if selection acts on a timescale that is substantially shorter than the timescale on which recombination and mutation events could negate the benefits.

There are a number of potential selective advantages given by co-transcription. In the case of operons that code for multi-protein complexes, co-transcription enables co-translational folding [6], it limits the half-life of toxic monomers [2], and it reduces stochastic differences in gene expression [8]. Operons that do not code for interacting proteins may be advantageous because of the co-regulation of protein expression [3]. Many examples of this class of operons are associated with metabolic operons [7] where co-regulated expression is likely to optimize the flux and to facilitate the regulation of functions, especially if these are required only under certain environmental conditions, or if complex regulatory structures are employed [3].

Evidence in favour of any of these proposed driving forces has so far largely been obtained from comparative genomics. Here we use a previously derived quantitative model for the network that controls the transcription factor σ^F during sporulation in *Bacillus subtilis* [4] to quantify the benefits of co-expression. Spore formation in *Bacillus subtilis* is a response to nutrient deprivation at high cell density and involves asymmetric septation and compartment-

specific initiation of gene expression [9]. The different gene programs in the larger mother cell and the smaller prespore are both directed by the transcription factor σ^F which, although only active in the smaller prespore, affects the transcriptional programs across the septum also in the mother cell, a phenomenon that is referred to as criss-cross regulation [10]. Successful sporulation therefore requires the rapid septation-dependent and prespore-specific activation of σ^F. σ^F is kept inactive by binding to SpoIIAB and is released upon binding of SpoIIAA (Fig. 1). SpoIIAA is phosphorylated by SpoIIAB [11] and reactivated by the serine phosphatase SpoIIE [12]. The balance between kinase and phosphatase activity thus determines whether or not σ^F is released from its inactive complex with SpoIIAB. SpoIIE accumulates on both sides of the asymmetrically positioned septum and therefore has an increased activity in the smaller compartment [13]. A quantitative model of the regulatory network predicts that because of the low turn-over rate most SpoIIE is bound by its substrate such that enzyme and substrate increase together in the smaller compartment [4]. According to the model, this combined increase is sufficient to trigger the formation of micromolar concentrations of σ^F holoenzyme in the prespore.

It is obvious from the above that the protein concentration ratio is important. An excess of σ^F or SpoIIAA compared to SpoIIAB will result in free σ^F and σ^F-dependent gene expression while an excess of SpoIIAB will prevent SpoIIAA-dependent σ^F release. In the vegetative cell the sporulation proteins are not detectable, and septation is preceded by $90 - 120'$ of gene expression, dependent on the exact experimental conditions [14, 15, 16]. Limiting the stochastic noise inherent in protein expression can be expected to be crucial for avoiding variations in the relative protein concentrations and the resulting sporulation defects. Three of the four proteins in the network are transcribed from genes in the *spoIIA* operon (Fig. 2A). These genes are not only co-transcribed into a single mRNA but are also most likely to be co-expressed since the translation of the three proteins appears to be coupled, at least to some degree. This system therefore offers an excellent opportunity to analyse the influence of transcriptional and translational co-regulation of the sporulation genes on an individual's survival, fitness.

Coupled translation is achieved when two genes are translated by the same ribosome. Reinitiation of translation at a nearby start codon after termination at the upstream gene is possible because ribosome dissociation from the mRNA is a slow and energy-dependent process [17]. There is currently no direct experimental evidence for coupled translation of the *spoIIA* operon. Such coupling can, however, be postulated based on the arrangement of genes [18]. The first two genes in the *spoIIA* operon (encoding SpoIIAA and SpoIIAB) overlap by four basepairs, while the genes for SpoIIAB and σ^F are interspaced by 11 basepairs (Fig. 2A); coupled translation has been documented for intercistronic distances of more than 60 basepairs [17]. The majority of genes that are organized in operons are separated by distances comparable to those found in the *spoIIA* operon [19], so that the studied system can be considered as representative of operons in general. The efficiency of reinitiation depends on the distance as well as the strength of the Shine-Dalgarno sequence [17, 20] which is, in general, located 5-13 basepairs upstream of a start codon and which binds to the homologous 3' end of the 16S rRNA, a component of the 30S ribosomal subunit. Moreover, the secondary structure of the mRNA can affect lateral diffusion of the ribosomes [20].

According to the protein expression data for the *spoIIA* operon it appears that the last gene in the operon, σ^F, is expressed at much lower levels than are SpoIIAA and SpoIIAB, while SpoIIAB monomers may be expressed at equal or up to 3-times higher levels compared to SpoIIAA [14, 15, 16]. The weaker expression of a downstream gene (as is the case for σ^F) can, in general, be accounted for by a weaker ribosomal binding site which is removed far enough from the termination codon of the upstream cistron that a considerable fraction of ribosomes dissociate from the mRNA before translation can be reinitiated [17]. It should be noted that while the transcriptional and translational coupling will reduce the noise in the relative SpoIIAB to σ^F expression levels the unbinding of ribosomes is necessarily a stochastic process and will therefore add a (low level) of noise. The stronger expression of a downstream gene (as may be the case for SpoIIAB relative to SpoIIAA) can, in general, only be observed if a strong initiation sequence for the downstream gene is occluded by mRNA secondary structure which is melted by the ribosome that transcribes the upstream gene [17]. Such a condition does not seem to be met by the gene for SpoIIAB, and more accurate expression data will be necessary to establish whether more SpoIIAB than SpoIIAA is expressed.

Available expression data can best be captured by an expression rate for SpoIIAB dimers and SpoIIAA of $6 \times 10^{-9} Ms^{-1}$ and of $2 \times 10^{-9} Ms^{-1}$ for σ^F and SpoIIE [4]; it should be noted that the simulation yields qualitatively similar results if SpoIIAB monomers and SpoIIAA are expressed at equal rates ($6 \times 10^{-9} Ms^{-1}$), as long as the σ^F and SpoIIE expression rate is then reduced to $10^{-9} Ms^{-1}$ [21]. As discussed in [21] the linear increase in the protein concentration assumed here does not fully match the experimental observations. There are, nonetheless, two good reasons to use a linear model. First of all, the data is too inaccurate and, in parts, contradictory to be modeled exactly. Secondly, the chosen rates correspond to the protein concentrations measured at the time of septation [14, 15, 16], the critical time point to judge sporulation success. This is because in the cell the IIE concentration increases more slowly than the other protein concentrations and only increases sharply immediately before septation [22]. As a consequence, the greatest danger of spontaneous uncompartmentalized activation of σ^F is just before septation, and this risk is fully assessed by the linear expression model. Since our analysis focuses mainly at what happens minutes before and after septation, individual fluctuations in the global expression rates during the 2 hours preceding septation are not important and the linear protein expression rates used should be considered as an averaged protein expression rate per bacterium.

Our quantitative ordinary differential equation model is very detailed - it comprises 50 dependent variables and 150 kinetic constants to describe the dynamics of only four proteins; the reader is referred to a detailed discussion of the model in the Supplementary Material of [4]. Given its high level of detail and accuracy the model predicts the phenotypes of essentially all mutants for which the biochemical effect is known. We can therefore expect that the predicted sporulation efficiencies in response to changes in parameter values are realistic. In the following we employ the model to quantify how far different levels of stochastic noise in gene expression, as modulated by different degrees of coupling of protein expression (that is by the coupling of both transcription and translation), affect the sporulation efficiency, that is the survival chances.

2 Results and Discussion

In the following we address how variations in the protein expression rates affect the sporulation efficiency. Here we will look at the effect of parallel changes in all protein expression rates as well as at the effects of independent changes that skew the ratios of protein concentrations. As the standard,"wild-type" protein expression rates we use $6 \times 10^{-9} Ms^{-1}$ for SpoIIAA and SpoIIAB dimers and $2 \times 10^{-9} Ms^{-1}$ for σ^F and SpoIIE [4]. After 120 minutes of protein expression the septum forms and SpoIIE accumulates on both sides of this septum. This is modeled by a four-fold increase in the concentration of SpoIIE, together with its associated substrate (phosphorylated SpoIIAA) in the prespore. As before we define a successful sporulation event by the requirement that before septation the concentration of $\sigma^F \cdot$ RNA polymerase holoenzyme does not exceed 0.4 μM while after septation the concentration exceeds one micromolar [4].

If the protein expression rates are all varied in parallel, that is by a common factor as denoted on the horizontal axis in Figure 2B, we find that the predicted sporulation efficiency is not affected as long as a minimal expression rate is kept to provide sufficient σ^F for binding to the RNA polymerase (Fig. 2B, grey lines). If the expression of SpoIIE is kept constant (in order to reflect that this protein is transcribed from a different locus and may therefore vary independently) then an independent 2.5-fold increase in the other sporulation proteins can still be tolerated before the relative activity of the phosphatase becomes too weak (Fig. 2B, black lines). An even higher independent increase in the expression of the *spoIIA* genes can be tolerated if we assume that the expression of the *spoIIA* and *spoIIE* genes is at least weakly correlated such that a large increase in the expression of the *spoIIA* genes is accompanied by a small increase in the expression of the *spoIIE* genes (Fig. 2C). Such a correlation is not unexpected considering that variations in gene expression are the result of both intrinsic and extrinsic noise. The latter, which reflects cell-to-cell variation in the concentration of other molecular species such as the RNA polymerase, will affect all genes similarly. We can conclude that the independent regulation of the *spoIIA* and *spoIIE* genes is unlikely to generate a major risk of failed sporulation. Separation of the *spoIIA* and *spoIIE* genes on the bacterial chromosome, on the other hand, has benefits because it ensures that, upon septation, each compartment retains one copy of *spoIIE* while initially (for the first $10 - 15'$) two copies of *spoIIA* are in the mother cell but none in the prespore [23]. This initial transient genetic imbalance may protect the mother cell from a relative increase of *spoIIE* to *spoIIA* gene products [21].

If the expression levels of the genes in the *spoIIA* operon are varied independently of each other, the tolerance of the network to variations in gene expression drops substantially. In particular, if SpoIIAB and SpoIIAA are no longer co-regulated, the network is sensitive to rather small changes (Fig. 2D, grey lines and circles). Thus if the SpoIIAA expression rate remains fixed and the SpoIIAB expression rate increases by 60% (corresponding to the factor 1.6 on the horizontal axis in Fig. 2D), then sporulation is predicted to fail; 60% variation from the mean is a noise level observed in bacterial (*E. coli*) expression systems [24]. On the other hand, if SpoIIAA and SpoIIAB remain co-regulated but σ^F expression is regulated independently (Fig. 2D, black lines), the network is rather robust to variations in gene expression as long as the

expression of SpoIIAB is increased more than the expression of σ^F and the overall σ^F concentration remains high enough to form micromolar concentrations of the holoenzyme. The transcriptional coupling together with a strong translational coupling of SpoIIAA and SpoIIAB therefore substantially increases the robustness of the network to fluctuations in gene expression. Stochastic variations in the relative rate of σ^F translation, on the other hand, is not as detrimental as long as the translation efficiency for σ^F is lower than for SpoIIAA and SpoIIAB, as can be achieved by a weaker ribosomal binding site and the resulting (stochastic) dissociation of ribosomes. An advantage of preferential dissociation of the ribosomes before translating the gene for σ^F is that the bacterium saves the energy that would otherwise be required to translate, and subsequently degrade, unnecessary (harmful) copies of σ^F. Considering that σ^F comprises 255 amino acids and linkage of each amino acid requires the equivalent of 4 ATPs the energy by not translating and degrading 10 μM σ^F corresponds to more than 10 mM ATP, which is a considerable amount considering that the bacterial ATP concentration is 1-3 mM [25, 26, 27] and sporulation is a response to starvation, that is energy deprivation.

In a last step we can now quantify the impact of gene organisation on sporulation efficiency, and therefore fitness. For this we assume that the gene expression levels in the cell population follow a normal distribution with variance η around the mean value. Given the complex regulation pattern of gene expression, gene expression levels are unlikely to be distributed exactly normally. A normal distribution is, however, still likely to provide an approximation no worse than what could be obtained with a detailed model of the regulatory process in the absence of sufficient data to determine all required parameter values [8]. Sporulation efficiency is determined as the fraction of simulation runs for which the concentration of $\sigma^F\cdot$ RNA polymerase holoenzyme does not exceed 0.4 μM before septation and exceeds one micromolar after septation [4]. For each condition the mean sporulation efficiency and standard deviation are calculated from 100 independent runs that are carried out 10 times. In each run the protein expression rates were set randomly such that overall the respective distributions of the protein expression rates were obtained. Determination of the sporulation efficiency for $\eta \in [0, 1]$ shows that as long as the sporulation genes are translationally coupled, even high variances hardly affect the sporulation efficiency (Fig. 3A, black lines). The sporulation efficiency is even higher at high noise level, η, if *spoIIE* expression co-varies with *spoIIA* expression, at least weakly (Fig. 3B). A lengthening of the transcription time, (that is a delay in septation) when transcription levels are too low to generate sufficient σ^F until septation will further increase robustness to fluctuations in the rate of protein expression. Such a dependency of the time point of septation on the protein (and in particular the SpoIIE) concentration is in agreement with experiments [28, 29] and might explain the large variance in the delay between the onset of sporulation and septation that is observed under different sporulation conditions. Low levels of additional stochastic noise in σ^F expression (broken lines), as may arise because of the stochastic dissociation of ribosomes, also has rather little impact and confirms that the weak coupling of SpoIIAB and σ^F translation does not substantially reduce sporulation efficiency. If on the other hand *spoIIAB* is removed from the operon and controlled independently by the same promotor then the sporulation efficiency drops rapidly (Fig. 3A, blue lines). This is in good quantitative agreement with experiments

which find that the sporulation efficiency drops to $40 - 80\%$ of wildtype levels [30], especially when considering that $\eta \sim [0.3, 0.6]$ for these expression levels [24]. If *spoIIAA* is moved instead, then the effect is reduced (Joanna Clarkson, personal communication) as also predicted by the model (Fig. 3A, grey lines).

It should be noted that this drop in sporulation efficiency has previously been accounted for by the loss of the transient genetic imbalance when *spoIIAB* is moved to a chromosomal position close to the origin of replication [30]. The transient lack of SpoIIAB expression in the prespore together with accelerated degradation of unbound SpoIIAB [31] had been suggested to enable σ^F release [30]. However, we have shown previously that the transient genetic imbalance does not affect σ^F release on the timescale on which it persists [21], and stochastic effects are therefore a much more likely explanation for the observed phenotype of the mutants.

We conclude from the analysis of this well studied model system that the protection from stochastic variation in the expression rate of interacting proteins can substantially increase viability, and therefore constitutes a driving force for gene clustering and co-regulation. Whilst the importance of gene dosage had been recognized before [32], and underexpression and overexpression of protein complex subunits in yeast had been shown to lower fitness [33], this study reveals that much smaller variances, as can result from stochastic effects, can already have substantial detrimental effects. The detailed analysis of the expression of the sporulation proteins therefore demonstrates the optimized character of gene regulation and suggests that co-regulation of genes serves to optimize cellular network dynamics in spite of the inherent noise in all biological processes.

Acknowledgements I wish to thank Iain D. Campbell, Joanna Clarkson, and Michael D. Yudkin for many valuable discussions, and Iain D. Campbell for his critical reading of the manuscript. The work was supported by a DTA EPSRC studentship as well as by a Junior Research fellowship held at St John's College, University of Oxford.

References

[1] Lawrence JG (1997) Selfish operons and speciation by gene transfer. *Trends Microbiol* **5**:355–359.

[2] Pal C, Hurst LD (2004) Evidence against the selfish operon theory. *Trends Genet* **20**:232–234.

[3] Price MN, Huang KH, Arkin AP, Alm EJ (2005) Operon formation is driven by co-regulation and not by horizontal gene transfer. *Genome Res* **15**:809–819.

[4] Iber D, Clarkson J, Yudkin MD, Campbell ID (2006) The mechanism of cell differentiation in *Bacillus subtilis*. *Nature* **441**:371–374.

[5] Jacob F, Monod J (1961) On the regulation of gene activity. *Cold Spring Harbor Symp Quant Biol* **26**:193–211.

[6] Dandekar T, Snel B, Huynen M, Bork P (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci* **23**:324–328.

[7] Lawrence JG, Roth JR (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* **143**:1843–1860.

[8] Swain PS (2004) Efficient attenuation of stochasticity in gene expression through post-transcriptional control. *J Mol Biol* **344**:965–976.

[9] Hilbert DW, Piggot PJ (2004) Compartmentalization of gene expression during *Bacillus subtilis* spore formation. *Microbiol Mol Biol Rev* **68**:234–262.

[10] Losick R, Stragier P (1992) Crisscross regulation of cell-type-specific gene expression during development in B. subtilis. *Nature* **355**:601–604.

[11] Min KT, Hilditch CM, Diederich B, Errington J, Yudkin MD (1993) σ^F, the first compartment-specific transcription factor of *B. subtilis*, is regulated by an anti-sigma factor that is also a protein kinase. *Cell* **74**:735–742.

[12] Duncan L, Alper S, Arigoni F, Losick R, Stragier P (1995) Activation of cell-specific transcription by a serine phosphatase at the site of asymmetric division. *Science* **270**:641–644.

[13] Arigoni F, Pogliano K, Webb C, Stragier P, Losick R (1995) Localization of protein implicated in establishment of cell type to sites of asymmetric division. *Science* **270**:637–640.

[14] Magnin T, Lord M, Yudkin MD (1997) Contribution of partner switching and SpoIIAA cycling to regulation of σ^F activity in sporulating *Bacillus subtilis*. *J Bacteriol* **179**:3922–3927.

[15] Lucet I, Borriss R, Yudkin MD (1999) Purification, kinetic properties, and intracellular concentration of SpoIIE, an integral membrane protein that regulates sporulation in *Bacillus subtilis*. *J Bacteriol* **181**:3242–3245.

[16] Lord M, Barillà D, Yudkin MD (1999) Replacement of vegetative σ^A by sporulation-specific σ^F as a component of the RNA polymerase holoenzyme in sporulating *Bacillus subtilis*. *J Bacteriol* **181**:2346–2350.

[17] McCarthy JE, Gualerzi C (1990) Translational control of prokaryotic gene expression. *Trends Genet* **6**:78–85.

[18] Fort P, Piggot PJ (1984) Nucleotide sequence of sporulation locus *spoIIA* in *Bacillus subtilis*. *J Gen Microbiol* **130**:2147–2153.

[19] Salgado H, Moreno-Hagelsieb G, Smith TF, Collado-Vides J (2000) Operons in Escherichia coli: genomic analyses and predictions. *Proc Natl Acad Sci U S A* **97**:6652–6657.

[20] Adhin MR, van Duin J (1990) Scanning model for translational reinitiation in eubacteria. *J Mol Biol* **213**:811–818.

[21] Iber D (2006) A computational analysis of the impact of the transient genetic imbalance on compartmentalized gene expression during sporulation in *Bacillus subtilis*. *J Mol Biol* **360**:15–20.

[22] Feucht A, Abbotts L, Errington J (2002) The cell differentiation protein SpoIIE contains a regulatory site that controls its phosphatase activity in response to asymmetric septation. *Mol Microbiol* **45**:1119–1130.

[23] Frandsen N, Barák I, Karmazyn-Campelli C, Stragier P (1999) Transient gene asymmetry during sporulation and establishment of cell specificity in *Bacillus subtilis*. *Genes Dev* **13**:394–399.

[24] Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* **297**:1183–1186.

[25] Jolliffe LK, Doyle RJ, Streips UN (1981) The energized membrane and cellular autolysis in *Bacillus subtilis*. *Cell* **25**:753–763.

[26] Guffanti AA, Clejan S, Falk LH, Hicks DB, Krulwich TA (1987) Isolation and characterization of uncoupler-resistant mutants of *Bacillus subtilis*. *J Bacteriol* **169**:4469–4478.

[27] Hecker M, Heim C, Volker U, Wolfel L (1988) Induction of stress proteins by sodium chloride treatment in *Bacillus subtilis*. *Arch Microbiol* **150**:564–566.

[28] Khvorova A, Zhang L, Higgins ML, Piggot PJ (1998) The spoIIE locus is involved in the Spo0A-dependent switch in the location of FtsZ rings in Bacillus subtilis. *J Bacteriol* **180**:1256–1260.

[29] Ben-Yehuda S, Losick R (2002) Asymmetric cell division in B. subtilis involves a spiral-like intermediate of the cytokinetic protein FtsZ. *Cell* **109**:257–266.

[30] Dworkin J, Losick R (2001) Differential gene expression governed by chromosomal spatial asymmetry. *Cell* **107**:339–346.

[31] Pan Q, Garsin DA, Losick R (2001) Self-reinforcing activation of a cell-specific transcription factor by proteolysis of an anti-sigma factor in *B. subtilis*. *Mol Cell* **8**:873–883.

[32] Veitia RA (2002) Exploring the etiology of haploinsufficiency. *Bioessays* **24**:175–184.

[33] Papp B, Pal C, Hurst LD (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**:194–197.

Figures

Figure 1: An overview of the interactions in the network that controls σ^F activity in *Bacillus subtilis*. For details see text. The figure is a reproduction of Figure 1 in [4].

Figure 2: The impact of parallel and random variations in the expression of *spoIIE* and *spoIIA* genes on σ^F release. **(A)** The *spoIIA* operon comprises the genes for SpoIIAA, SpoIIAB, and σ^F. The genes for SpoIIAA and SpoIIAB overlap; the genes for SpoIIAB and σ^F are separated by 11 basepairs. **(B)** The regulatory network is robust to parallel variations in gene expression. The predicted concentration of σ^F·RNApolymerase holoenzyme before (dashed lines) and after septum formation (continuous lines) if either all (grey lines) or all protein expression rates except for the one of SpoIIE (black lines) were increased by the factor on the horizontal axis compared to the standard reference rates ($6 \times 10^{-9} Ms^{-1}$ for SpoIIAA and SpoIIAB dimers and $2 \times 10^{-9} Ms^{-1}$ for σ^F and SpoIIE [4]). **(C,D)** The expression rate combinations for which septation-dependent σ^F release is possible (between the lines) or not (outside the area marked by lines). **(C)** The impact of differential regulation of *spoIIE* and *spoIIA* expression. The vertical and horizontal axes indicate the fold variation in the *spoIIE* and *spoIIA* expression rates respectively, compared to the standard reference rates. **(D)** The impact of differential regulation of the expression of genes encoded in the *spoIIA* operon. The vertical axis indicates the fold variation in the expression of SpoIIAA (circles), σ^F (black lines), or SpoIIAA and σ^F (grey lines). The horizontal axis indicates the fold variation in the expression of SpoIIAB and of any other protein whose expression is coupled to the one of SpoIIAB (which are those genes in the *spoIIA* operon not reported on the vertical axis). The sudden jump observed at a high SpoIIAB to σ^F ratio (lower black line) is the consequence of impaired σ^F release when the relative SpoIIAB concentration is too high.

Figure 3: The impact of stochastic variation in gene expression on sporulation efficiency. **(A)** The fraction of successful sporulation events (as defined in the text) dependent on the variance in gene expression if expression of the *spoIIA* genes is either coupled (black lines), the expression of SpoIIAB and σ^F is coupled (grey lines), or the expression of SpoIIAA and σ^F is coupled (blue lines). SpoIIE is expressed throughout at the standard rate of 2×10^{-9} M^{-1} s^{-1}. The broken lines show the effect of an additional independent normal variation in the rate of σ^F expression with $\eta_S = 0.1$ (dashed lines) or $\eta_S = 0.3$ (dotted lines) from the coupled rates. If σ^F is one of the coupled rates then σ^F expression is varied both together with its coupling partner and additionally independently to reflect the additive levels of noise acting at the initiation of translation and the re-initiation/dissociation step. **(B)** The fraction of successful sporulation events (as defined in the text) dependent on the variance in gene expression if expression of the *spoIIA* and *spoIIE* genes is coupled (to assess the benefits of correlated expression), and an additional noise term η_E is added to the expression of *spoIIE* with $\eta_E = 0.1$ (black continuous line), $\eta_E = 0.3$ (dotted line), or $\eta_E = 0.6$ (dashed line); η_E assesses the effects of independent

promotors and spatial heterogeneity in the concentration of transcription and translation factors. The red line is identical to the continuous black line in panel A (noise in coupled *spoIIA* expression, SpoIIE expressed at 2×10^{-9} M^{-1} s^{-1}). Mean and standard deviation are based on 10 times 100 independent runs.

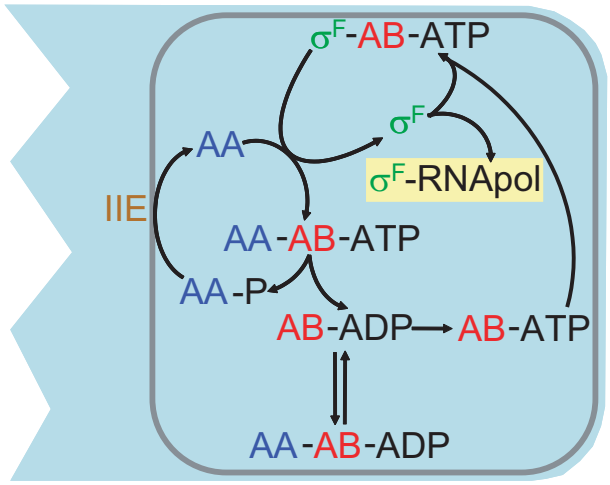

Figure 1:

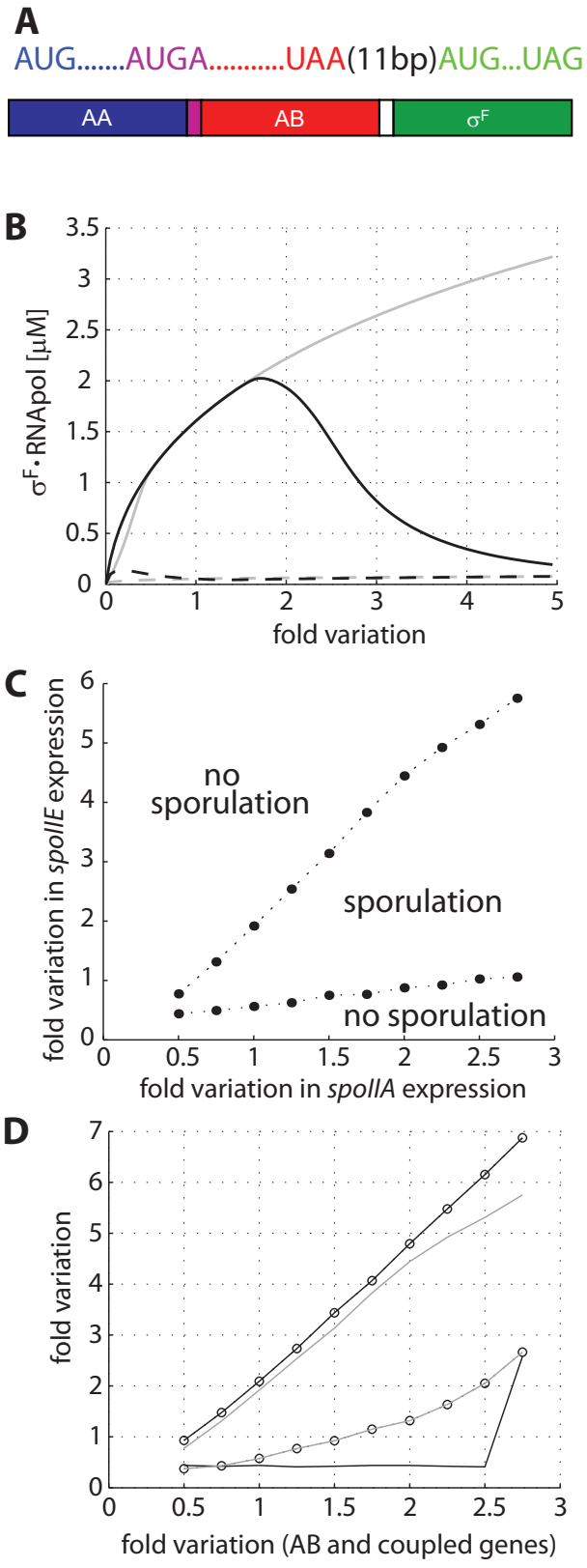
A
AUG.....AUGA.....UAA(11bp)AUG...UAG
AA
AB
σF
B
σF• RNApol [μM]
fold variation
C
fold variation in spoIIE expression
no sporulation
sporulation
no sporulation
fold variation in spoIIA expression
D
fold variation
fold variation (AB and coupled genes)

Figure 2:

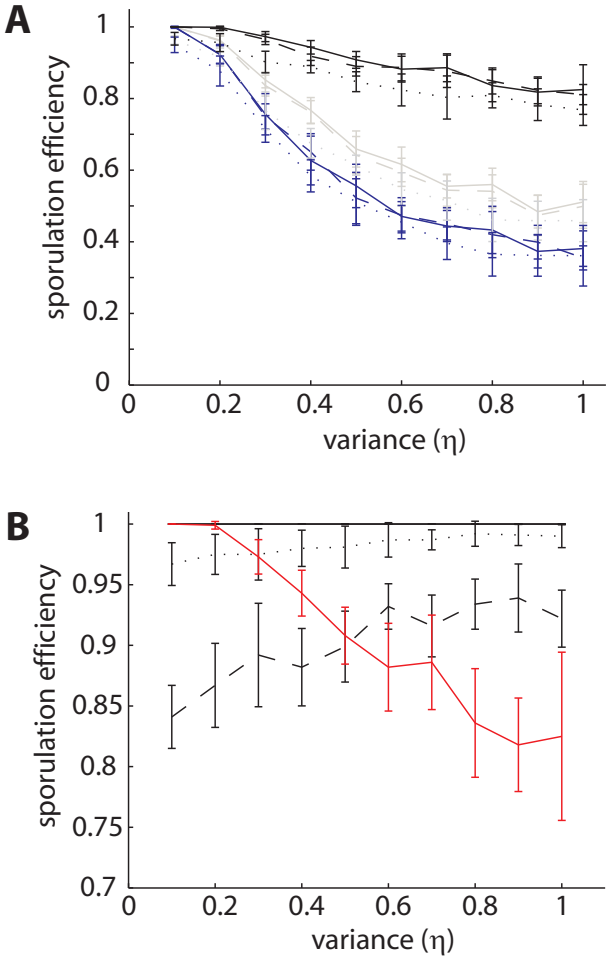
A
sporulation efficiency
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
variance (η)
B
sporulation efficiency
1
0.95
0.9
0.85
0.8
0.75
0.7
0
0.2
0.4
0.6
0.8
1
variance (η)

Figure 3: